# Lec 12 MATH 343

HW: prove $H-H_1$ is idempotent

We proved last time using

$H_1 := \frac{1}{n}J_n$, $H = X(X^\top X)^{-1}X^\top$ that

$$\frac{\dfrac{\|\hat{\vec{y}} - X\beta - \bar{y}\vec{1}_n + H_1 X\vec{\beta}\|^2}{p}}{\dfrac{\|\vec{e}\|^2}{n-(p+1)}} \sim F_{p,\,n-(p+1)}$$

this allows testing of any $H_0: \vec{\beta} = \vec{\theta}$

However, we are usually interested in the omnibus/global test of $H_0: \beta_1 = \beta_2 = \dots = \beta_p = 0$ which means none of the features have any linear predictive power.

Analyzing the numerator under $H_0$

$$\left\|\hat{\vec{y}} - X\begin{bmatrix}\beta_0\\0\\\vdots\\0\end{bmatrix} - \bar{y}\vec{1}_n + H_1 X\begin{bmatrix}\beta_0\\0\\\vdots\\0\end{bmatrix}\right\|^2 = \|\hat{\vec{y}} - (\beta_0\vec{1}_n) - \bar{y}\vec{1}_n + H_1(\beta_0\vec{1}_n)\|^2$$

$$= \|\hat{\vec{y}} - \cancel{\beta_0\vec{1}_n} - \bar{y}\vec{1}_n + \cancel{\beta_0\vec{1}_n}\|^2 = \|\hat{\vec{y}} - \bar{y}\vec{1}_n\|^2.$$

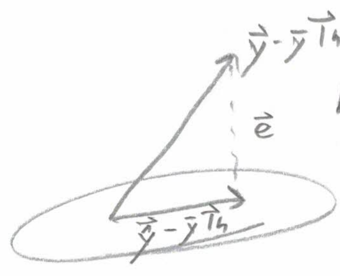
The estimator under $H_0$ is:

let $MSR := \frac{SSR}{p}$

$$\hat{F} = \frac{\dfrac{\|\hat{\vec{y}} - \bar{y}\vec{1}_n\|^2}{p}}{\dfrac{\|\vec{e}\|^2}{n-(p+1)}} \sim F_{p,\,n-(p+1)}.$$

The test statistic is $\hat{\hat{F}} = \dfrac{\frac{SSR}{p}}{\frac{SSE}{n-(p+1)}} = \dfrac{MSR}{MSE}$

$p\text{-val} = P(\hat{F} > \hat{\hat{F}})$




ratio of average levels of signal per feature to average levels of error per observation

Relationship between ominbus F statistic and $R^2$.

$$\hat{F} = \frac{\overbrace{n-(p+1)}^{q}}{p} \frac{SSR}{SSE} \Rightarrow \frac{q}{p} \frac{SSR}{SST-SSR} \Rightarrow \frac{1}{\hat{F}} = \frac{p}{q} \frac{SST-SSR}{SSR}$$

$$\Rightarrow \frac{1}{\hat{F}} = \frac{p}{q}\left(\frac{SST}{SSR}-1\right) \Rightarrow \frac{1}{\hat{F}} = \frac{p}{q}\left(\frac{1}{R^2}-1\right) \Rightarrow \frac{1}{\hat{F}} = \frac{p}{qR^2} - \frac{p}{q} \Rightarrow \frac{p}{qR^2} = \frac{1}{\hat{F}} + \frac{p}{q}$$

$$\Rightarrow \frac{1}{R^2} = \frac{q}{p\hat{F}} + 1 \Rightarrow \frac{1}{R^2} = \frac{q+p\hat{F}}{p\hat{F}} \Rightarrow R^2 = \frac{p\hat{F}}{n-(p+1)+p\hat{F}} \in (0,1)$$

$\hat{F} \to 0 \Leftrightarrow R^2 \to 0$

$\hat{F} \to 1 \Leftrightarrow R^2 \to 1$

$p+1 = n \Rightarrow R^2 = 1$

MATH 309

transform of variable

$$\Longrightarrow f_{R^2}(R^2) = \text{Beta}\left(\frac{p}{2}, \frac{n-(p+1)}{2}\right)$$

Cool fact!

Let's play this game ~~again~~.

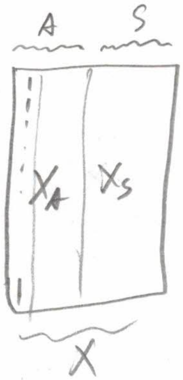
Let $S \subset \underbrace{\{1,2,\dots,p\}}_{\text{indices of features}}$, let $A = \underbrace{\{0\}}_{\text{index of intercept column}} \cup \underbrace{S^c}_{\text{the rest of the features}}$

Let $\vec{\beta}_S$ be the $\beta_j$'s s.t. $j \in S$

Let $\vec{\beta}_A$ be the $\beta_j$'s s.t. $j \in A$

where $k := |S| \Rightarrow |A| = p+1-k$

Now, rearrange columns of of design matrix so that:




Let $H_A = X_A(X_A^T X_A)^{-1} X_A^T$ be the proj. matrix onto colsp$(X_A)$

Note $H = (H_A) + (H - H_A)$

$\vec{y} = H\vec{y} = (H_A + (H - H_A))\vec{y} = H_A\vec{y} + (H - H_A)\vec{y}$ Now show they're orthogonal

$(H_A\vec{y})^T (H - H_A)\vec{y} = \vec{y}^T H_A^T (H\vec{y} - H_A\vec{y}) = \vec{y}^T H_A H \vec{y} - \vec{y}^T H_A^T H_A \vec{y}$

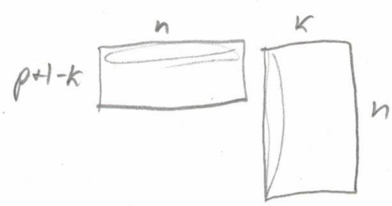
$= \vec{y}^T H_A H \vec{y} - \|H_A\vec{y}\|^2 = \vec{y}^T H_A \vec{y} - \|H_A\vec{y}\|^2 = \vec{y}^T H_A^T H_A \vec{y} - \|H_A\vec{y}\|^2 = \|H_A\vec{y}\|^2 - \|H_A\vec{y}\|^2 = 0$

$H_A H \vec{y} = \text{proj}_{X_A}(\text{proj}_X(\vec{y})) \overset{\text{intuitively}}{=\!=} \text{proj}_{X_A}(\vec{y}) = H_A\vec{y} \Rightarrow H_A H = H_A$

Proof. Let $X_A = Q_A R_A$, $X = QR$. Note $Q = [Q_A \mid Q_{-A}]$ by Gram-Schmidt Algorithm

$H_A H = (Q_A Q_A^T)(QQ^T) = Q_A Q_A^T [Q_A \mid Q_{-A}] \begin{bmatrix} Q_A^T \\ Q_{-A}^T \end{bmatrix} = Q_A Q_A^T (Q_A Q_A^T + Q_{-A} Q_{-A}^T)$

$= Q_A \underbrace{\cancel{Q_A^T Q_A}}_{I_{p+1-k}} Q_A^T + Q_A \underbrace{Q_A^T Q_{-A}}\, Q_{-A}^T = Q_A Q_A^T + O_{n,n} = Q_A Q_A^T = H_A$




since each are orth. by G-S, $Q_A^T Q_{-A} = O_{p+1-k,\,k}$

rank$[H_A] = p+1-k$

rank$[H - H_A] = k$

$$\frac{1}{\sigma^2}\vec{\varepsilon}^T\vec{\varepsilon} = \frac{1}{\sigma^2}\vec{\varepsilon}^T H_A \vec{\varepsilon} + \frac{1}{\sigma^2}\vec{\varepsilon}^T(H-H_A)\vec{\varepsilon} + \frac{1}{\sigma^2}\vec{\varepsilon}^T(I-H)\vec{\varepsilon}$$

by Cochran's

$$\Longrightarrow \frac{1}{\sigma^2}\vec{\varepsilon}^T H_A \vec{\varepsilon} \sim \chi^2_{p+1-k} \text{ indep of } \frac{1}{\sigma^2}\vec{\varepsilon}^T(H-H_A)\vec{\varepsilon} \sim \chi^2_k \text{ indep of } \frac{1}{\sigma^2}\vec{\varepsilon}^T(I-H)\vec{\varepsilon} \sim \chi^2_{n-(p+1)}$$

$$\frac{1}{\sigma^2}\vec{\varepsilon}^T(H-H_A)\vec{\varepsilon} = \frac{1}{\sigma^2}\vec{\varepsilon}^T(H-H_A)^T(H-H_A)\vec{\varepsilon} = \frac{1}{\sigma^2}\|(H-H_A)\vec{\varepsilon}\|^2 = \frac{1}{\sigma^2}\|(H-H_A)(\vec{Y}-X\vec{\beta})\|^2$$

$$= \frac{1}{\sigma^2}\|HY - H_A\vec{Y} - HX\vec{\beta} + H_AX\vec{\beta}\|^2 = \frac{1}{\sigma^2}\|\hat{\vec{y}} - \hat{\vec{y}}_A - X\beta + H_AX\vec{\beta}\|^2$$

$$= \frac{1}{\sigma^2}\left\|\hat{\vec{y}} - \hat{\vec{y}}_A - [X_A\ X_S]\begin{bmatrix}\vec{\beta}_A\\ \vec{\beta}_S\end{bmatrix} + H_A[X_A\ X_S]\begin{bmatrix}\vec{\beta}_A\\ \vec{\beta}_S\end{bmatrix}\right\|^2 = \frac{1}{\sigma^2}\|\hat{\vec{y}} - \hat{\vec{y}}_A - \cancel{X_A\vec{\beta}_A} - X_S\vec{\beta}_S + \underbrace{\cancel{H_AX_A\vec{\beta}_A}}_{X_A\vec{\beta}_A} + H_AX_S\vec{\beta}_S\|^2$$

$$= \frac{1}{\sigma^2}\|\hat{\vec{y}} - \hat{\vec{y}}_A + (I-H_A)X_S\vec{\beta}_S\|^2$$

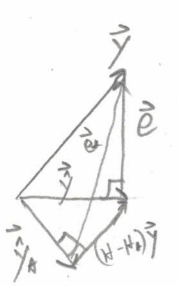
$$\Longrightarrow \frac{\frac{\frac{1}{\sigma^2}\vec{\varepsilon}^T(H-H_A)\vec{\varepsilon}}{k}}{\frac{\frac{1}{\sigma^2}\vec{\varepsilon}^T(I-H)\vec{\varepsilon}}{n-(p+1)}} \sim F_{k,n-(p+1)} \Rightarrow \frac{\frac{\|\hat{\vec{y}} - \hat{\vec{y}}_A + (I-H_A)X_S\vec{\beta}_S\|^2}{k}}{\frac{\|\vec{e}\|^2}{n-(p+1)}} \sim F_{k,n-(p+1)}$$

This can be used to run ~~tests~~ of $H_0: \vec{\beta}_S \neq \vec{\theta}$

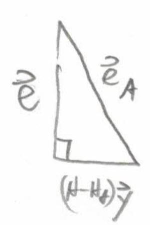
Typically, the test of interest is $H_0: \vec{\beta}_S \neq \vec{0}_K$

Using this $H_0$, the estimator is:

$$\hat{F} = \frac{\frac{\|\hat{\vec{y}} - \hat{\vec{y}}_A\|^2}{k}}{\frac{\|\vec{e}\|^2}{n-(p+1)}} = \frac{\frac{\|\vec{e}_A\|^2 - \|\vec{e}\|^2}{k}}{\frac{\|\vec{e}\|^2}{n-(p+1)}} \sim F_{k,n-(p+1)}$$




$\Rightarrow$




$$\Rightarrow \|\vec{e}\|^2 + \|(H-H_A)\vec{y}\|^2 = \|\vec{e}_A\|^2 \Rightarrow \|(H-H_A)\vec{y}\|^2 = \|\vec{e}_A\|^2 - \|\vec{e}\|^2$$

(Proof using algebra on HW)

The test statistic is then:

$$\hat{\hat{F}} = \frac{\overbrace{SSE_A - SSE}^{\Delta SSE \leftarrow \text{ is the addition of the } k \text{ features ``worth it''?}}\big/ k}{MSE}, \quad pval = P\left(\hat{F} > \hat{\hat{F}}\right)$$

$SSE_A$ is sometimes called SSE of the "reduced model" as it has a reduced # of features ($p+1-k$) relative to the "full model" which has all $p+1$ features. Note that the reduced model is nested in the full model.